California State Water Resources Control Board

Optional Deliverable

Trial format version to be evaluated over three months

GeoTracker

Field Data Collection Guidelines & Restrictions

Electronic Deliverable Format and Data Dictionary

September 8, 2004

Web site: http://geotracker.swrcb.ca.gov/

Table of Contents

List of Tables

Acronyms

CSV	Comma Separated Values (also known as comma/quote delimited)
EDD	Electronic Data Deliverable
EDF	Electronic Deliverable Format
GIS	Geographic Information System
LUST	Leaking Underground Storage Tank
PK	Primary Key
RP	Responsible Party
SWRCB	(California) State Water Resources Control Board
UST	Underground Storage Tank
VVL	Valid Value List

1 Introduction

AB2886 (Water Code Sections 13195-13198) requires Responsible Parties to electronically submit compliance data, such as soil or water chemistry analysis, location, and elevation data to the SWRCB Geographical Environmental Information Management System (GeoTracker). In accordance with emergency regulations adopted by the SWRCB (Article 12, Chapter 16, Division 3, Title 23 of the California Code of Regulations), electronic reporting of data will be in addition to the submittal of hard copy (paper) reports such as site investigations and quarterly monitoring reports to the lead agency. Electronic reporting of monitoring well locations and data related to monitoring well elevation data would be required as of January 1, 2002.

The GeoTracker Field Data Collection electronic data deliverable (EDD) consists of field sampling information and information allowing that information to be blinded to the laborator. This deliverable is intended to be produced by the Responsible Party (or their consultant) and to be imported into the GeoTracker system via Internet. The GeoTracker Field Data Collection data, when uploaded into the SWRCB GeoTracker database, will provide essential information pertaining to groundwater and soil sampling points.

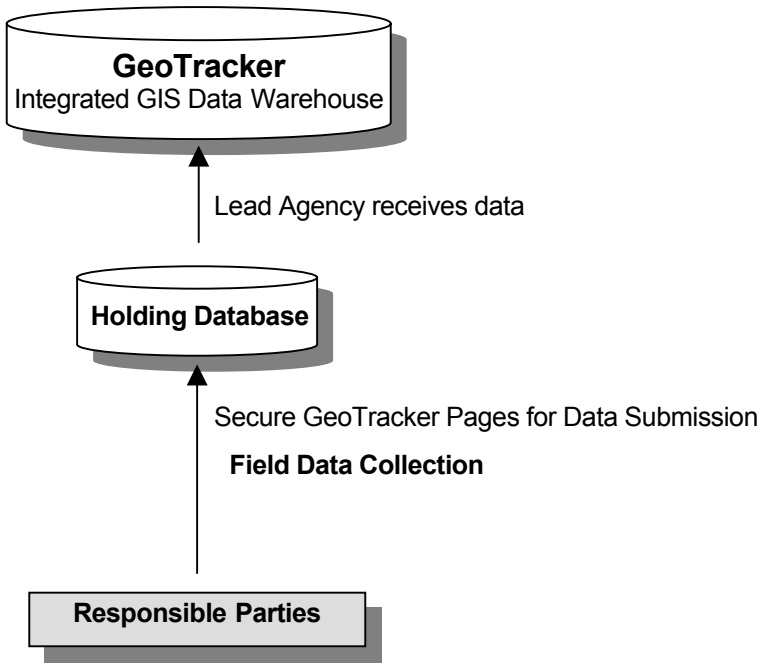


Figure 1: Data Flow for the GeoTracker Field Data Collection Deliverable

1.1 Document Conventions

This document presents the structure and guidelines and restrictions for creating a GeoTracker Field Data Collection electonic data deliverable. Each data file of the deliverable is discussed in a level of detail that assists the user in creating an electronic deliverable that meets the criteria of the data standard. Included is a discussion of guidelines and restrictions that apply to files and those that apply to individual fields.

1.1.1 Figure Representation of Files

Each file discussion begins with a figure representing the fields in the file. Refer to Figure 2 as an example. The fields are listed in the order in which they exist within the structure, and primary key fields are underlined. "Primary key" means a selected field (or fields in combination) that makes a record unique in a database. Refer to the Glossary in Appendix B for a technical definition of this and other terms. The order of the fields in the figure **is** the order expected for delivery.

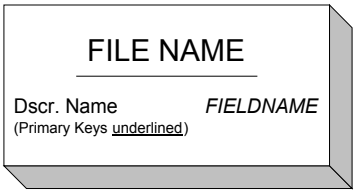


Figure 2: Example Figure Definition

1.1.2 Table Representation of GeoTracker Field Data Collection Files

The following table is a representation of the tables defining files of the GeoTracker Field Data Collection.

Field Name	Attrb	VVL	REQ	Dscr. Name	Definition	Guidelines & Restrictions
FIELD1	C18	Yes	Yes	Field 1		Field 1 is a character field with 18 available positions.
FIELD2	D10	No	Yes	Field 2		Field 2 is a date field with an expected format of MM/DD/YYYY.
FIELD3	N5	No	Yes	Field 3		Field 3 is a numeric field with a total of 5 spaces available for numbers and decimals, with no restriction on the number of digits to the right of the decimal point other than the overall field size.

The "Field Name" is the actual structural name of the field. All primary key fields are shaded within these tables (e.g., *FIELD1* and *FIELD2*). Primary key fields are identified within the tables by shading. **Fields are listed in their structural order within these tables**.

"Attrb" describes the field attributes (type and size). For example:

- C18 is an 18-character field (alphanumeric).
- D10 is a date field with an expected format of MM/DD/YYYY (i.e., 01/01/2001).
- N5 is a numeric field with a total of 5 spaces available for numbers and decimals, with no restriction on the number of digits to the right of the decimal point other than the overall field size (e.g., 12345 or 123.4 or 1.234).

The "VVL" column indicates with a "Yes" or "No" whether the data field requires a valid value code.

The "REQ" column indicates with a "Yes" or "No" whether entry into a field is required.

The "Dscr. Name" column gives the descriptive name of the field.

The "Definition" is a brief definition and/or explanation of the field.

The "Guidelines & Restrictions" describe expectations for entry into the field.

1.1.3 Conventions for Text

Throughout this document, file names are capitalized (e.g., the GEO_FIELD file), and field names are capitalized and italicized (e.g., the *GLOBAL_ID* field). The words "file" and "table" are used interchangeably. The description of each file includes such information as which fields require valid values, which fields require entry for submission, and the file's primary keys.

1.2 Valid Values (VVLs)

Various data fields in the GeoTracker Field Data Collection electronic deliverable require entry of valid values. Valid values are built-in codes that the format requires for certain fields, such as datum and survey method. The reason for using specific values for these fields is to standardize the data entry, to ensure data consistency, and to help prevent errors. Freely entered data might contain extra spaces, commas, or dashes that would make meaningful data manipulation and thorough or accurate data searches impossible.

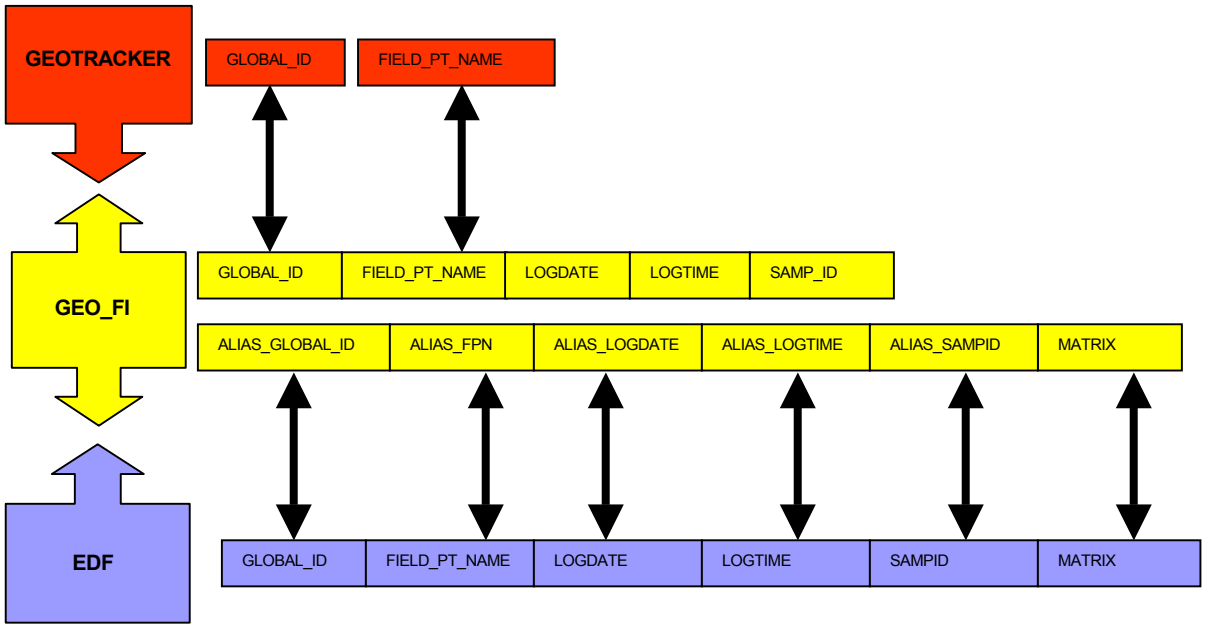
Most valid values are abbreviations of common or proper names; hence selecting the correct code is generally straightforward. A list of codes is provided for all fields requiring valid values within each file format sections 3.1.3, 3.2.3, and 3.3.3.

2 Data Deliverable Description

The GeoTracker system relies on a relational database consisting of files related to one another through common (key) fields. These data files are described as relational because the information in one file is related to information in other files, linked through a group of fields called the primary key. The primary key fields in one file record must be identical to the same fields in the linking file record in order to "relate" the data records in both files.

The GEO_FIELD file contains information pertaining to the sampling information from the field and additional fields that allow the blinding of that information to the analytical laboratory. The GEO_FIELD relates to information in GeoTracker through the primary keys fields of *GLOBAL_ID* and *FIELD_PT_NAME*. The FIELD_CS relates to the EDF analytical data by *ALIAS_FPN, ALIAS_SAMP_ID, ALIAS_LOGTIME, and ALIAS_GLOBAL_ID.*

Figure 3. Relational Structure for GeoTracker Field Data Collection files



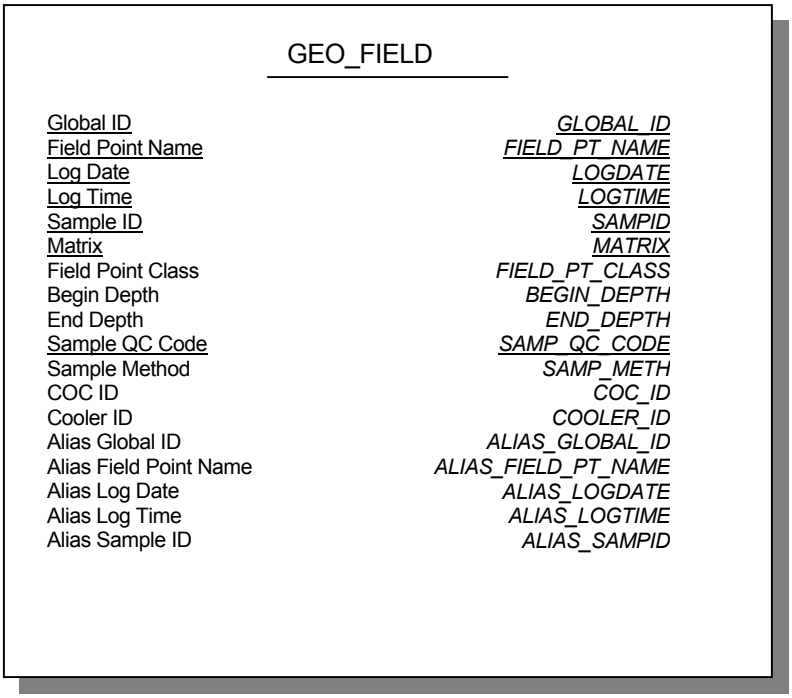
3 Relational Files Format

The following Chapter describes the format, data dictionary and guidelines and restrictions associated with each of the data files of the GeoTracker Field Data Collection electronic deliverable.

3.1 GEO_FIELD: The Field Collection Data file.

The GEO_FIELD provides field collection data for import into the GeoTracker system. The GEO_FIELD additionally gives the ability to blind field sampling information from the analytical laboratory.

-



3.1.1 File Guidelines and Restrictions:

- Primary key fields: *GLOBAL_ID, FIELD_PT_NAME, LOGDATE, LOGTIME, SAMP_ID, SAMP_QC_CODE, and MATRIX* comprise the primary keys.
- Required fields: *GLOBAL_ID, FIELD_PT_NAME, LOGDATE, LOGTIME, SAMP_ID, MATRIX, FIELD_PT_CLASS, BEGIN_DEPTH, SAMP_QC_CODE, ALIAS_GLOBAL_ID, ALIAS_FIELD_PT_NAME, ALIAS_LOGDATE, ALIAS_LOGTIME, and ALIAS_SAMP_ID* require entry.
- Valid Value fields: *MATRIX, FIELD_PT_CLASS*, *SAMP_QC_CODE*, and *SAMPLE_METHOD* require valid value entries.

3.1.2 The GEO_FIELD File Format

Table 1: GEO_FIELD Format

Field Name	Attrb	VVL	REQ	Dscr. Name	Definition	Guidelines & Restrictions
GLOBAL_ID	C12	Yes	Yes	Global ID Number	The unique identifier for a regulated facility or site.	Obtain from GeoTracker web pages.
FIELD_PT_ NAME	C15	Yes	Yes	Field Point Name	The field name or common name of the location where the field sample has been collected (i.e. Well 01)	Obtain from GeoTracker web pages. Add new field point names via GeoTracker AB2886 electronic reporting interface.
LOGDATE	D8	No	Yes	Collection Date	The date a field sample is collected.	Date format is MM/DD/YYYY.
LOGTIME	C4	No	Yes	Collection Time	The time that a field sample is collected, recorded using 24-hour military time	Time format is HHMM.
SAMPID	C20	No	Yes	COC Sample ID	The sample identification as listed on the Chain-of-Custody. In many cases this will be identical to the Field Point Name.	
MATRIX	C2	Yes	Yes	Matrix	The code identifying the sample matrix as determined by the laboratory (e.g., water, soil, etc.)	
FIELD_PT_CLASS	C5	Yes	Yes	Class of field point	Code describing the type of "point" being sampled.	
BEGIN_DEPTH	C7	No	Yes	Sample beginning depth	Beginning depth of sample – depth in feet from ground surface at which a sample is collected.	
END_DEPTH	C7	No	No	Sample ending depth	End depth of sample – depth in feet from ground surface that defines the lower boundary of the sampling interval (i.e. soil samples).	
SAMP_QC_CODE	C5	Yes	Yes	Type of	Code that describes if the sample is a specific	

				sample	type of QC sample, duplicate or spilt sample.	
SAMP_METHOD	C6	No	No	Sampling method	Code that describes the sampling method used.	
COC_ID	C12	No	No	COC ID	Unique ID that links to the chain of custody	
COOLER_ID	C25	No	No	Cooler ID	ID of the cooler used to transport the sample	
ALIAS_GLOBAL_ID	C12	No	No	Alias global ID	The blinded unique identifier for a regulated facility or site.	This couple be generated or be the same as the GLOBAL_ID.
ALIAS_FPN	C15	No	No	Alias Field Point Name	The blinded field name or common name of the location where the field sample has been collected (i.e. Well 01)	This couple be generated or be the same as the FIELD_PT_NAME.
ALIAS_SAMPID	C20	No	No	Alias sample ID	The blinded sample identification as listed on the Chain-of-Custody. In many cases this will be identical to the Field Point Name.	This couple be generated or be the same as the SAMPID.
ALIAS_LOGDATE	C4	No	No	Alias sample collection time	The blinded time that a field sample is collected, recorded using 24-hour military time	Date format is MM/DD/YYYY.
ALIAS_LOGTIME	C4	No	No	Alias sample collection time	The blinded time that a field sample is collected, recorded using 24-hour military time	Time format is HHMM.

3.1.3 Valid Value Lists for GEO_FIELD

Table 2: *FIELD_PT_CLASS* valid values

Code	Description
AGIR	Agriculture/irrigation well
AGT	Above Ground Tank Location
AMB	Ambient drinking water aquifer monitoring well (NOT a remediation well)
BH	Borehole
BM	Benchmark
CENT	Location of facility/site surveyed at the centroid of UST field
CS	Composite Sample
FAGT	Former Above Ground Tank Location
FL	Fuel Line Sampling Point
FUST	Former Underground Storage Tank, Tank Pit Location
MW	Remediation/ Groundwater Monitoring well
PRIW	Domestic/Private Drinking Water Well
PUBW	Public Drinking Water Well
RES	Lake or Reservoir Sampling Point
RIV	River Sampling Point
SP	Spring
SS	Stockpile Sample
STR	Stream or Creek Sampling Point
SUR	Surface Sampling Point
TRS	Transient Subsurface Sampling Point (i.e. geoprobe…)
UST	Underground Storage Tank, Tank Pit Location
VEW	Vapor Extraction Well
WSFI	Water System Facility Intake

Table 3: *SAMP_QC_CODE* valid values

Note: **The number (#) next to each code should be filled in starting with 1 through 9.

Code	Description
AB#	Ambient Conditions Blank
FB#	Field Blank
FD#	Field Duplicate
FR#	Field Replicate

Code	Description
FS#	Field Spike
CS	Normal Client Sample
TB#	Trip Blank

Table 4: *MATRIX* valid values

Code	Description
GS	Soil Gas
WG	Ground Water
SO	Soil

Table 5: *SAMP_METH* valid values

Code	Description
AC	Air Canister
GB	Geoprobe
CC	Core Sampler
SS	Split Spoon
WF	Wellhead Faucet (Grab Sample From)
SL	Suction Lift Pump

4.0 Submittal File, Record, and Data Field Requirements

It is mandatory that the file, record, and data field requirements identified below be adhered to in order to generate acceptable upload file.

4.1 Submittal File Requirements

An EDD may be submitted as a fixed length or tab-delimited format including: ASCII *.TXT file, Microsoft Excel™ tab delimited *.XLS file, or comma separated value (CSV) delimited ASCII *.TXT file (also known as "comma/quote delimited").

Each line of data is equivalent to a single record in the data submission. Each record is made up of distinct fields of information. Listed below are the file and record specifications for entering each record of data in its specified file.

- The column heading or field name is not part of the file and should be omitted. Only the data being entered should be in the upload file. Inclusion of column headers (i.e. the field names) in the upload document will cause errors. (Specific instructions for use of Microsoft Excel™ and Microsoft Access™ are detailed below).
- If entering the data via a spreadsheet, such as Microsoft Excel™, enter the first field's data in the A1 cell. Each cell following A1 should be an additional field of data. When the document is saved as tab-delimited text, excel will save it in the proper format for entry (see instructions below).
- If the record is being created in a text-editor program, there should be no margins and no word-wrapping. (Text-editor programs include Notepad™ and Wordpad™). In each file, every record starts in the farthest left position of "position number 1."
- Empty rows (records) are not allowed. The first record or row in the file, and every subsequent record or row, must contain valid data.
- Every record within a file must be unique. If, for each key field, a record's data appears exactly the same in another record, these two records are considered to be duplicate records.

4.2 Submittal Data Field Requirements

When producing the fixed or tab delimited formats, data element formats (attributes) must be strictly followed.

- Only authorized codes from the valid value list should be keyed into fields requiring valid values.
- Valid data must always be entered for all required fields.
- Do not add or delete any fields.
- Optional fields where data is not being reported should be left blank (will be converted to unknown).

4.2.1 Tab Delimited Formats

Creating a tab delimited file from a Microsoft Excel™ spreadsheet:
In order to create a file for submittal, create an Excel™ spreadsheet with column headings that match the fields and order required for the GEO_FIELD file (specified in this document). Enter valid data for each LUST case in a single row, with each field occupying one cell. Enter the first field's data in the A1 cell. Each cell following A1 should be an additional field of data. Save the spreadsheet as a text document (tab delimited), which is one of the options under file>save as> save as type. At this point, Excel™ will generate a message warning that some formatting may be lost by saving your document in this format. Simply click the "Yes" button, and save the file. By saving as "text (tab delimited)", Excel™ omits the column headers and margins from the document as required for upload. This is the "formatting" that Excel™ warns will be lost when saving the document. This file can be uploaded through the GeoTracker Web-based system.

Creating a tab delimited file from Microsoft Access™
In order to create a file for submittal, first create a database consistent with the GEO_FIELD file guidelines. To export data as a tab delimited file, go to file>Export, and set the "Save as type" as "Text Files" and press the "Save All" button. In the new window that will appear, select the "Delimited" option, press "Next", and choose tab delimited. Press "Next" again, and then press "Finish". The new tab-delimited text document will be saved in the directory that was chosen during export.

4.2.2 Comma Delimited Format

CSV delimited format:
For the CSV delimited format, the value entered must be equal to or shorter than the field's length and followed by a comma. For example, when entering a *GLOBAL_ID*, which is a C12 field, if the value to be entered is only C5, in the CSV delimited format it would look like:

"12345","next field entry"

4.2.3 Fixed Length Format

The fixed length format:
Data fields in a file are limited to a certain number of spaces and the data must be in a specific position. Character data must be left justified within a field. Numeric data must be right justified within a field. If the information to be entered is shorter than the field width, insert blank spaces in the field's remaining positions. If the data to be entered is longer than the allowed field width, the data must be shortened to a unique identifier or significant value.

The start- and end-position numbers indicate the exact character locations where the applicable data must be placed in the file. There are some cases where the field is a single character wide. It, therefore, has the same start- and end-position number. The single character of data must be put in that position of the record. For example, when entering a *GLOBAL_ID*, which is a C12 field, if the value to be entered is only C5, in the fixed length format, it would look like:

12345…….next field entry

(where the dots represents 7 blank spaces before the next field).

4.3 Procedure for EDD Submittal

Files may be error checked and submitted via the Internet to the SWRCB GeoTracker system. Dynamic error reports will be generated to inform the user of any data format issues that must be resolved. The files must pass the error-checking program before the data will be submitted to the lead regulatory agency and the GeoTracker database.

Appendix A: Glossary of Terms

Attributes - The format and size attributes of a database field. A field type of C8 is a field that can hold up to eight alphanumeric characters. An N5 field type has a total of 5 spaces available for numbers and decimals, with no restriction on the number of digits to the right of the decimal point other than the overall field size (e.g., 12345 or 123.4 or 1.234). A D10 field type is a date field, and is formatted as MM/DD/YYYY ([month]/[day]/[year]).

Database - A collection of information arranged into records (rows) and fields (columns) for ease of sorting and manipulation within a table or related tables.

Deliverable - A report, data, etc., that is "delivered" to another party, either electronically, or in hard copy format.

EDD (Electronic Data Deliverable) - Information stored in a defined format, accessible via a computer (e.g., stored on diskette, internal hard drive, CD ROM, magnetic tape, etc.).

Field - An area of a table (a column) that contains a particular piece of information. One or more fields make a record. Fields are defined by the attributes of format and size.

File - A named group of electronic data in a defined format.

Guidelines and Restrictions - Information provided to the user regarding data entry, data performance, and data delivery expectations.

Primary Key - A field or set of fields that uniquely identify a record within a table. Key fields within a table define the primary key. Each database record can be uniquely identified using the combination of data fields that make up the primary key.

Record - A line of data (a row) in a table or file made up of distinct fields of information.

Responsible Party- The individual or organization legally responsible for the assessment, monitoring and/or remediation of a contaminated site.

Table - A format for data that allows for data manipulation within a database. Tables are organized with columns and rows of information.

Valid Value - Specially assigned, standardized coded value designating an approved (i.e., "valid") value for entry into a field in the database.